Estado de la publicación: No informado por el autor que envía

LatArXiv. Creación de un repositorio preprint y datos de investigación abierto para la comunidad latinoamericana

Patricio Iván Pantaleo

https://doi.org/10.1590/SciELOPreprints.8124

Envíado en: 2024-06-27
Postado en: 2024-07-05 (versión 1)
(AAAA-MM-DD)

SciELO Preprints - Este documento es un preprint y su situación actual está disponible en: https://doi.org/10.1590/SciELOPreprints.8124

LatArXiv. Creación de un repositorio preprint y datos de investigación abierto para la comunidad latinoamericana.

Patricio Iván Pantaleo
Paideia Studio – Universidad Pablo de Olavide
https://orcid.org/0000-0002-8104-8975

RESUMEN. LatArXiv (https://preprints.latarxiv.org/) es una iniciativa para crear un repositorio preprint y de datos de investigación abierta específicamente para la comunidad latinoamericana. Busca relevar y potenciar las mejores prácticas de repositorios existentes, enfocándose en el contexto de Latinoamérica, aunque también está abierto a contribuciones del ámbito iberoamericano y otras regiones, siempre con un enfoque en Latinoamérica.
En este trabajo se ha llevado a cabo un relevamiento y análisis de repositorios de preprints y datos de investigación de importancia a nivel mundial. Para ello, se utilizaron plantillas estandarizadas REBIUN para evaluar características como estructura, políticas de acceso abierto y metadatos. Asimismo, se empleó SemRush, una herramienta de SEO, para analizar el rendimiento en línea de los repositorios, considerando indicadores como el Authority Score y el tráfico orgánico.
Los resultados revelaron que, si bien existen repositorios con un compromiso significativo con la visibilidad y accesibilidad de los preprints y datos, aún hay áreas de mejora, como la presencia en directorios y recolectores, mejora en la interfaz del usuario y la claridad de las políticas. Se identificaron correlaciones entre el cumplimiento de criterios REBIUN y el rendimiento SEO de los repositorios, destacando la importancia de cumplir con estándares de calidad para mejorar la visibilidad en línea.
Basándose en estos hallazgos, se desarrolló un Producto Mínimo Viable (MVP) en preprints.latarxiv.org, con el objetivo de fomentar la visibilidad y accesibilidad de la investigación en la comunidad científica hispanohablante. El MVP se diseñó priorizando una interfaz intuitiva, políticas claras y estrategias de promoción activa. Además, se establecieron canales de retroalimentación para mejorar continuamente la plataforma. Aunque el análisis de los repositorios de datos está en una etapa inicial, se observaron tendencias similares, resaltando la importancia de mantener altos estándares de calidad y promover la visibilidad en línea de los datos de investigación en el ámbito hispanohablante.
Este trabajo se presentó como trabajo final del Diplomado en Experto de Servicios de Apoyo a la Investigación: Información, Datos, Evaluación y Edición Científica de la Universidad Pablo de Olavide de España (2024).

PALABRAS CLAVES. LatArXiv, repositorio preprint, datos de investigación, comunidad latinoamericana, OPS, DSpace, Dataverse, repositorios abiertos, repositorios latinoamericanos, SEO, bibliometría, Iberoamérica.

SciELO Preprints - Este documento es un preprint y su situación actual está disponible en: https://doi.org/10.1590/SciELOPreprints.8124

LatArXiv. Creation of an open preprint and research data repository for the Latin American community.

Patricio Iván Pantaleo
Paideia Studio – Universidad Pablo de Olavide
https://orcid.org/0000-0002-8104-8975

ABSTRACT. LatArXiv (https://preprints.latarxiv.org/) is an initiative to create a preprint and open research data repository specifically for the Latin American community. It seeks to emulate the best practices of existing repositories, focusing on the Latin American context, although it is also open to contributions from Latin America and other regions, always with a focus on Latin America.
In this work, a survey and analysis of repositories of preprints and research data of worldwide importance has been carried out. Standardized REBIUN templates were used to evaluate characteristics such as structure, open access policies and metadata. SemRush, an SEO tool, was also used to analyze the online performance of the repositories, considering indicators such as Authority Score and organic traffic.
The results revealed that while there are repositories with a significant commitment to the visibility and accessibility of preprints and data, there are still areas for improvement, such as presence in directories and harvesters, improved user interface and clarity of policies. Correlations were identified between compliance with REBIUN criteria and SEO performance of repositories, highlighting the importance of complying with quality standards to improve online visibility.
Based on these findings, a Minimum Viable Product (MVP) was developed at preprints.latarxiv.org, with the objective of promoting the visibility and accessibility of research in the Spanish-speaking scientific community. The MVP was designed prioritizing an intuitive interface, clear policies and active promotion strategies. In addition, feedback channels were established to continuously improve the platform.
Although the analysis of data repositories is at an early stage, similar trends were observed, highlighting the importance of maintaining high quality standards and promoting the online visibility of research data in the Spanish-speaking world.
This paper was presented as the final work of the Diploma in Expert in Research Support Services: Information, Data, Evaluation and Scientific Editing of the Pablo de Olavide University of Spain (2024).

KEYWORDS. LatArXiv, preprint repository, research data, Latin American community, OPS, DSpace, Dataverse, open repositories, Latin American repositories, SEO, bibliometrics, Ibero-America.

SciELO Preprints - Este documento es un preprint y su situación actual está disponible en: https://doi.org/10.1590/SciELOPreprints.8124

Introducción

En el vertiginoso mundo de la investigación científica, la rápida difusión de resultados es fundamental para impulsar el progreso del conocimiento. En este contexto, los repositorios de preprints y datos juegan un papel crucial al proporcionar plataformas para compartir hallazgos de manera temprana y transparente. Sin embargo, la selección adecuada de estos repositorios es esencial tanto para autores como para editores a fin de garantizar visibilidad y una correcta difusión del contenido académico.

El proyecto de LatArXiv se centra en analizar repositorios de acceso abierto en América Latina y España, tanto para PrePrints como para Set de Datos con el fin de sintetizar las mejores prácticas y oportunidades a la hora de implementar una nueva propuesta de repositorios. Para lograr este objetivo, es fundamental comprender el panorama actual de estos repositorios y evaluar su rendimiento en términos de visibilidad en línea, políticas de acceso abierto, calidad de los recursos e interoperabilidad. De esta manera, se avanza en el estudio de una selección de repositorios y en la realización de un producto en base a este relevamiento. Tanto el estudio como el desarrollo tiene una división fundamental entre repositorios preprint y repositorios de datos; sin embargo, el desarrollo del producto actual se ha iniciado por el repositorio preprint quedando para una etapa posterior el desarrollo del repositorio de datos.

En la sección de Materiales y Métodos, se describe detalladamente el proceso de selección de los repositorios a relevar, así como las herramientas y metodologías utilizadas para recopilar y analizar datos. Se aborda desde la identificación de repositorios relevantes hasta la evaluación de su desempeño en línea a través de herramientas como las plantillas REBIUN y SemRush.

A través de este análisis exhaustivo, se busca identificar áreas de mejora y recomendar acciones específicas para fortalecer la infraestructura de acceso abierto en la región. Este enfoque metodológico riguroso proporciona una base sólida para la toma de decisiones informadas y la implementación de mejoras efectivas en el proyecto LatArXiv.

La propuesta de trabajo final para LatArXiv surge como respuesta a la necesidad imperante de contar con un repositorio preprint y de datos de investigación abierto que atienda específicamente las demandas y particularidades de la comunidad de habla hispana. Este proyecto se inspira en el éxito de iniciativas como SciELO Preprints y SciELO Data, adaptando sus mejores prácticas y proponiendo mejoras.

Estado de la cuestión

LatArXiv se estructura en dos grandes componentes: el repositorio de preprints, basado en OPS (Open Preprint Systems), y el repositorio de datos, basado en plataformas como DSpace o Dataverse. Esta estructura permite ofrecer un espacio centralizado y accesible para compartir tanto resultados de investigación preliminares como conjuntos de datos relevantes.

Una de las características distintivas de LatArXiv es su apertura a cualquier autor o editorial que desee depositar preprints o datos de investigación, sin restricciones de afiliación o membresía. Esto amplía el alcance y la diversidad del contenido disponible

SciELO Preprints - Este documento es un preprint y su situación actual está disponible en: https://doi.org/10.1590/SciELOPreprints.8124

en el repositorio, promoviendo la colaboración y el intercambio de conocimientos en la región.

En el contexto de los preprints, LatArXiv implementa un sistema de secciones que permite a las revistas o editoriales asociadas tener su propia URL para depósitos, facilitando la identificación y el acceso al contenido. Además, se incorpora el sistema InReview, que indica el estado del preprint en el proceso de revisión por pares, brindando transparencia y visibilidad a este proceso.

Para los datos de investigación, LatArXiv ofrece estructuras flexibles basadas en DSpace o Dataverse, permitiendo tanto depósitos independientes como espacios específicos para revistas o editoriales adheridas. Se prioriza el desarrollo de políticas éticas claras, términos y condiciones transparentes, y medidas de privacidad acordes a las normativas internacionales para garantizar la integridad y confianza en el repositorio.

Actualmente, LatArXiv se encuentra en fase de desarrollo inicial, con un MVP disponible solo para preprints en https://preprints.latarxiv.org/. Este MVP presenta una plantilla CSS moderna, integración con herramientas como Google Analytics y Mailchimp, y una funcionalidad inicial del sistema InReview. El financiamiento inicial se sustenta en un sistema de membresía, con publicaciones gratuitas para los autores, aunque este modelo puede evolucionar según las necesidades y avances del proyecto. También se ha implementado un foro (https://forum.latarxiv.org/) para la comunidad basado en un reconocido software de discusión como Discourse y se prevé desarrollar a futuro el repositorio de datos (https://data.latarxiv.org/).

En resumen, LatArXiv se presenta como una propuesta innovadora y necesaria para fortalecer la infraestructura de acceso abierto en América Latina, proporcionando un espacio inclusivo y transparente para compartir y acceder a la investigación científica en la región. A continuación, se detalla la selección y análisis realizados que posibilitaron conocer el contexto y desarrollar el primer producto para LatArXiv.

Materiales y Métodos

Selección de Repositorios a Relevar

Para comprender mejor la relevancia de estos repositorios de preprints y de datos, es fundamental contextualizar el papel que desempeñan en el ecosistema de la investigación científica actual. Con el avance de la tecnología y la globalización, se ha vuelto cada vez más importante la rápida difusión de los resultados de investigación para promover la colaboración, la retroalimentación y el avance del conocimiento en todas las áreas del saber.

Esto, sin embargo, no carece de críticas que fortalecen y visualizan los puntos a mejorar por nuevas propuestas; a la vez que los preprints y otros repositorios se convierten en estadíos previos que validan lo difundido en revistas científicas o se convierten en *cuellos de botellas* fundamentales para pasos posteriores en la comunicación científica (Giles, 2003; Kaiser, 2017; Kwon, 2020; van Schalkwyk & Dudek, 2022).

Frente a esto, delimitar los repositorios referentes para analizar es un desafío que se abordó en dos etapas:

SciELO Preprints - Este documento es un preprint y su situación actual está disponible en: https://doi.org/10.1590/SciELOPreprints.8124

1. En la primera etapa, se accedió a los datos disponibles en OpenDOAR, un directorio global de repositorios de acceso abierto, para obtener información sobre los repositorios relevantes para el proyecto LatArXiv. La robusta API y la buena documentación que tiene OpenDOAR facilitó una segmentación y relevamiento inicial utilizando librerías Pyhton.
   Se accedió a la base de datos de OpenDOAR para recopilar información sobre los repositorios disponibles en América Latina y España. Utilizando la interfaz de OpenDOAR, se realizaron búsquedas específicas utilizando criterios como la región geográfica, el tipo de repositorio (preprints y repositorios de datos), y áreas temáticas relevantes para la comunidad científica latinoamericana.

   Los datos obtenidos incluyen información sobre el nombre del repositorio, la URL, la institución responsable, la tecnología utilizada y otras características de importancia. Una vez recopilados los datos de OpenDOAR, se aplicaron criterios de selección y filtrado para identificar los repositorios más relevantes para el proyecto LatArXiv. Estos criterios incluyeron la calidad del repositorio, la relevancia para la comunidad científica latinoamericana, la disponibilidad de datos abiertos, la interoperabilidad y otros aspectos importantes para el proyecto.

   Esta primera etapa dio como resultado un contexto amplio y fundamentado en datos y accesibilidad que en gran parte está documentado y es replicable desde este repositorio de GitHub: https://github.com/Ppantaleo/LatArXiv-context

   Sin embargo, cercenar el relevamiento a una sola base de datos, aunque muy funcional, dejaba elementos sin definir, principalmente en lo que respecta a preprints. Los resultados eran muy generalistas incluyendo entre los seleccionados a recolectores como LA Referencia y nos impedía avanzar con el análisis sobre un único portal. De la misma forma, sucedía con DIALNET en la segmentación de DataSets. Por estos motivos, se avanzó en una segunda etapa.

2. La primera etapa sentó las bases para un relevamiento secundario que avanzó en la inclusión de dos bases de datos adicionales: https://doapr.coar-repositories.org/repositories/ y https://www.re3data.org/. La inclusión de estas bases de datos adicionales en el relevamiento secundario se justifica por la necesidad de abordar limitaciones identificadas en la primera etapa del análisis y obtener una visión más completa del panorama de los repositorios de acceso abierto útiles para analizar el desarrollo de una propuesta para América Latina y España.
   Las nuevas fuentes proporcionan conjuntos de datos complementarios y permiten una segmentación más detallada de los repositorios de preprints y datos de investigación. La inclusión de estas bases de datos adicionales amplía la muestra de repositorios considerados y facilita la identificación de repositorios que podrían haber sido pasados por alto en la primera etapa del análisis. Como se dijo, estas fuentes ofrecen una oportunidad para realizar un relevamiento más detallado y específico, lo que contribuye a obtener una visión más completa y representativa del panorama de los repositorios de acceso abierto a nivel mundial y su impacto en la región.

En definitiva, si la primera etapa brindó acceso documentado a una base de datos amplia de repositorios, la segunda etapa posibilitó definir los repositorios específicos que se estudian aquí. Los criterios definitivos que se implementaron para definir el corpus de análisis siguieron estas variables cualitativas:

SciELO Preprints - Este documento es un preprint y su situación actual está disponible en: https://doi.org/10.1590/SciELOPreprints.8124

1. **Reputación y Renombre:** Considera la reputación y el renombre de los repositorios en la comunidad científica internacional. Los repositorios que tienen una larga trayectoria, una sólida reputación y son ampliamente reconocidos por los investigadores de todo el mundo son buenos candidatos para la inclusión en el análisis.
2. **Impacto en la Comunidad Científica:** Evalúa el impacto que tienen los repositorios en la comunidad científica en términos de la cantidad y calidad de los recursos que albergan, así como su contribución al avance del conocimiento en sus respectivas áreas temáticas.
3. **Calidad de los Recursos:** Analiza la calidad de los recursos disponibles en los repositorios, como la relevancia de los preprints y datos de investigación, la exhaustividad de la información proporcionada, y la disponibilidad de metadatos y documentación relevante.
4. **Interoperabilidad y Accesibilidad:** Considera la interoperabilidad de los repositorios y su capacidad para integrarse con otras plataformas y servicios. Además, evalúa la accesibilidad de los recursos para investigadores de diferentes regiones y disciplinas.
5. **Institucionalidad:** En el caso de los repositorios de datos, considera la institucionalidad del repositorio y la reputación de la institución responsable. Los repositorios respaldados por instituciones académicas o de investigación bien establecidas suelen ser más confiables y tienen mayor credibilidad.

Finalmente, se decidió por la selección de los siguientes 10 repositorios, 5 corresponden a PrePrints y 5 a repositorios de Datos:

PrePrint

En Latinoamérica no es posible identificar variedad de repositorios preprints, por lo que aquí se generaliza la selección. El carácter abierto es una característica de estos repositorios.

1. **arXiv (https://arxiv.org/):**

Desde su creación en 1991, arXiv ha sido pionero en la difusión de preprints en disciplinas como física, matemáticas, ciencias de la computación y áreas relacionadas. Su impacto en la comunidad científica es innegable, ya que ha facilitado el acceso temprano a investigaciones importantes, ha agilizado el proceso de revisión por pares y ha fomentado la colaboración entre investigadores de todo el mundo. Su larga trayectoria y su prestigio lo convierten en un punto de referencia para aquellos que buscan compartir y acceder a investigaciones de vanguardia.

2. **bioRxiv (https://www.biorxiv.org/):**

bioRxiv ha surgido como una plataforma líder en la difusión de preprints en el campo de la biología. Con un enfoque en disciplinas como biología molecular, genética, neurociencia y biología celular, bioRxiv ha democratizado el acceso a la investigación biológica y ha acelerado el ritmo de descubrimientos en estas áreas. Su crecimiento exponencial en los últimos años refleja la creciente demanda de compartir resultados de investigación de manera rápida y transparente en la comunidad científica.

3. **SciELO Preprints (https://preprints.scielo.org/):**

SciELO Preprints - Este documento es un preprint y su situación actual está disponible en: https://doi.org/10.1590/SciELOPreprints.8124

SciELO Preprints es parte del Sistema SciELO, una iniciativa que promueve la visibilidad y accesibilidad de la investigación científica en América Latina y el Caribe. Al centrarse en la región latinoamericana, SciELO Preprints ofrece una plataforma para compartir investigaciones que de otro modo podrían no recibir la atención que merecen en el ámbito internacional. Además de aumentar la visibilidad de la investigación latinoamericana, SciELO Preprints fomenta la colaboración entre investigadores de la región y contribuye al avance del conocimiento local e internacional.

4. **OSF Preprints (https://osf.io/preprints/):**

OSF Preprints es parte del Centro para la Ciencia Abierta (Center for Open Science, COS), una organización que aboga por la transparencia y la apertura en la investigación científica. Esta plataforma permite a los investigadores compartir preprints en una amplia variedad de disciplinas, promoviendo la colaboración interdisciplinaria y la transparencia en la comunicación científica. Su integración con otros servicios ofrecidos por COS, como la gestión de proyectos de investigación y la colaboración en línea, lo convierte en un recurso integral para la comunidad científica.

5. **ChinaXiv (http://chinaxiv.org/home.htm):**

ChinaXiv se ha convertido en un importante centro de difusión de investigación científica producida en China y otras regiones asiáticas. Al proporcionar una plataforma para compartir preprints en una amplia variedad de disciplinas, ChinaXiv contribuye al intercambio de conocimientos en la comunidad científica asiática y promueve la colaboración entre investigadores de la región y de todo el mundo. Su crecimiento y alcance reflejan el papel cada vez más importante que desempeña China en el panorama científico global. Junto con SciELO constituyen los repositorios con identificación geográfica, similar a la propuesta de LatArXiv, es por este motivo que se ha incluido en el relevamiento si bien su impacto es notablemente más bajo que los demás.

Datos

Para repositorios de Datos, fue más fácil identificar portales de Latinoamérica ya que son más abundantes que repositorios preprints. Por este motivo, todos los seleccionados son de la región. Sin embargo, la mayoría son propuestas institucionales y nos son abiertos a la comunidad científica en general.

1. **Repositorio Institucional de la Universidad Nacional de Colombia (repositorio.unal.edu.co)**:

Este repositorio proporciona un espacio centralizado para almacenar y compartir datos de investigación generados por la comunidad académica de la Universidad Nacional de Colombia. Al igual que los preprints, los datos de investigación son un componente fundamental de la comunicación científica y su acceso abierto promueve la transparencia, la reproducibilidad y el avance del conocimiento en diversas áreas del saber.

2. **Repositorio ANID (repositorio.anid.cl)**:

El Repositorio ANID, administrado por la Agencia Nacional de Investigación y Desarrollo (ANID) de Chile, cumple un papel crucial en la preservación y difusión de

SciELO Preprints - Este documento es un preprint y su situación actual está disponible en: https://doi.org/10.1590/SciELOPreprints.8124

los datos de investigación producidos en el país. Al igual que los preprints, los datos de investigación son una parte esencial del proceso científico y su acceso abierto facilita la colaboración, la verificación de resultados y el desarrollo de nuevas investigaciones.

3. **Sedici - Repositorio Institucional de la Universidad Nacional de La Plata (sedici.unlp.edu.ar)**:

Sedici es el repositorio institucional de la Universidad Nacional de La Plata en Argentina, que almacena una amplia variedad de materiales académicos y científicos, incluidos los datos de investigación. Al igual que los preprints, los datos de investigación son valiosos para la comunidad científica y su disponibilidad abierta en repositorios como Sedici promueve la colaboración, la innovación y el avance del conocimiento.

4. **Datos Abiertos UNAM (datosabiertos.unam.mx)**:

La Universidad Nacional Autónoma de México (UNAM) promueve la apertura y accesibilidad de los datos de investigación a través de su repositorio de datos abiertos. Al igual que los preprints, los datos de investigación son esenciales para la construcción de nuevos conocimientos y su disponibilidad abierta en repositorios como Datos Abiertos UNAM fomenta la transparencia y la colaboración en la comunidad científica.

5. **SciELO - Datos Abiertos (data.scielo.org)**:

SciELO es una plataforma que promueve la visibilidad y accesibilidad de la investigación científica en América Latina y el Caribe. A través de su sección de Datos Abiertos, SciELO proporciona un repositorio para compartir conjuntos de datos asociados a publicaciones científicas. Al igual que los preprints, los datos de investigación son esenciales para la validación y reproducción de resultados, y su acceso abierto en repositorios como SciELO contribuye al avance del conocimiento en la región.

Relevamiento y Análisis de Datos

Una vez definido el corpus de sitios de referencias para relevar datos se avanzó con dos herramientas básicas que se encuentran detalladas en la sección de ANEXOS:

1. **Plantillas REBIUN**: REBIUN es la Red de Bibliotecas Universitarias Españolas, y proporciona plantillas estandarizadas para la descripción y evaluación de repositorios institucionales (Barrueco Cruz, et al., 2021). Estas plantillas están diseñadas para recopilar información detallada sobre los repositorios, incluyendo aspectos como su estructura, políticas de acceso abierto, metadatos, interoperabilidad, estadísticas de uso, entre otros. Al utilizar estas plantillas, se busca estandarizar la información recopilada sobre los repositorios seleccionados, lo que facilita su comparación y análisis. Por lo tanto, en el contexto del proyecto LatArXiv, el uso de las plantillas REBIUN permite una evaluación detallada de sus características y funcionalidades.

SciELO Preprints - Este documento es un preprint y su situación actual está disponible en: https://doi.org/10.1590/SciELOPreprints.8124

Los resultados de la evaluación se dividen en dos secciones: "Cumple" y "No Cumple". En cada sección, se detallan los criterios específicos que son cumplidos o no por el repositorio, junto con una breve explicación o justificación. Esto proporciona una visión detallada de las fortalezas y áreas de mejora del repositorio en cuestión.

2. **Información SEO mediante SemRush**: SEO (Search Engine Optimization) se refiere a las prácticas destinadas a mejorar la visibilidad de un sitio web en los resultados de búsqueda orgánica de los motores de búsqueda. SemRush es una herramienta de análisis SEO que proporciona información detallada sobre el rendimiento de un sitio web en términos de tráfico, palabras clave, enlaces entrantes, entre otros aspectos. En el contexto del proyecto LatArXiv, el análisis de información SEO mediante SemRush podría implicar el uso de esta herramienta para evaluar el rendimiento y la visibilidad en línea de los repositorios seleccionados. Esto podría incluir la identificación de las palabras clave más relevantes para cada repositorio, el análisis de la competencia en términos de SEO, la evaluación de la calidad de los enlaces entrantes, entre otros aspectos. Este análisis proporciona información valiosa sobre la presencia en línea de los repositorios seleccionados y ayuda a identificar áreas de oportunidad para mejorar su visibilidad y alcance.

   Para esto, se utilizó una versión paga de SemRush que permite un análisis general de subdominio y/o directorios para lograr restringir el análisis directamente al repositorio, cuando estos se encuentran en un subdominio, sin contemplar variables del dominio que, para nuestro caso, no contribuyen al estudio específico. De cada sitio analizado se adjunta en ANEXO el PDF general del relevamiento del sitio en SemRush con diferentes índices para el último mes. En el análisis SEO se retoman los siguientes indicadores de SemRush y fueron tomados el 21 y 22 de abril de 2024:

   - **Authority Score**: Este indicador proporciona una medida de la autoridad o relevancia de un sitio web en comparación con otros sitios en su industria o nicho. Se basa en factores como la calidad de los enlaces entrantes y la calidad del contenido.
   - **Tráfico orgánico**: Se refiere a la cantidad de visitantes que llegan al sitio web a través de resultados de búsqueda orgánica en motores como Google, Bing, entre otros. Este indicador muestra la efectividad del SEO y la visibilidad en línea del sitio.
   - **Backlinks**: Los backlinks son enlaces entrantes desde otros sitios web que apuntan al sitio en cuestión. Este indicador muestra la cantidad y calidad de los enlaces que contribuyen a la autoridad y visibilidad del sitio.
   - **Palabras claves orgánicas**: Indica las palabras clave por las que el sitio web está clasificado en los resultados de búsqueda orgánica. Esto ayuda a entender qué términos de búsqueda son relevantes para el sitio y su contenido.
   - **Intención en palabras claves**: Este indicador analiza la intención detrás de las palabras clave utilizadas por los usuarios en sus búsquedas. Puede indicar si los usuarios están buscando información, productos, servicios, etc.

SciELO Preprints - Este documento es un preprint y su situación actual está disponible en: https://doi.org/10.1590/SciELOPreprints.8124

- **Competidores orgánicos**: Identifica a los competidores más relevantes del sitio web en términos de SEO, es decir, otros sitios que compiten por las mismas palabras clave y tráfico orgánico.

Finalmente, con el relevamiento realizado se analizaron los datos obtenidos siguiendo los siguientes parámetros:

- Análisis de los resultados de las plantillas REBIUN: Se evalúa el cumplimiento de los criterios definidos para cada repositorio en función de las plantillas REBIUN y se identifica patrones o tendencias comunes entre los repositorios en términos de visibilidad, políticas, metadatos, interoperabilidad, logs y estadísticas.
- Interpretación de los datos de SemRush: Se analiza los indicadores proporcionados por SemRush para cada repositorio y se compara el rendimiento de los repositorios entre sí y con respecto a la competencia en términos de SEO.
- Correlación entre los resultados de las plantillas REBIUN y SemRush: Se busca correlaciones entre los resultados obtenidos a través de las plantillas REBIUN y los datos de SemRush; por ejemplo, ¿existe una relación entre la calidad de los metadatos de un repositorio y su autoridad en línea medida por SemRush? Finalmente, se explora cómo los aspectos evaluados en las plantillas REBIUN pueden influir en el rendimiento SEO de los repositorios.
- Identificación de áreas de mejora y recomendaciones para LatArXiv: Basándonos en los resultados de las plantillas REBIUN y SemRush, se identifican áreas específicas en las que los repositorios pueden mejorar su funcionamiento y visibilidad en línea para realizar una propuesta e implementación en el proyecto de LatArXiv.

Resultados y Discusión

El proceso de relevamiento y análisis de repositorios para el proyecto LatArXiv ha proporcionado una visión detallada de la principal infraestructura disponible para la difusión de preprints y datos de investigación. Esto es una base de referencia ineludible para pensar el desarrollo de nuevas propuestas en el contexto de habla hispana.

Resultados Generales

La presencia de repositorios preprints en el ámbito de habla hispana se encuentra en un estadío inicial con gran potencial de desarrollo. Respecto a los repositorios de datos, la práctica institucional está más desarrollada. Tanto por el incentivo de programas nacionales y regionales, el depósito de datos, tesis u otros objetos en repositorios institucionales es una práctica más desarrollada que el depósito de preprints. Sin embargo, de igualmente manera que los servidores de prepublicaciones, es difícil concebir un repositorio de datos abierto naturalmente tanto en la recepción como en la difusión del contenido. SciELO ha llevado a cabo sus estrategias en esta vía, pero la difusión no es amplia más allá del ecosistema de revistas integradas a la red.

Finalmente, sobre la metodología implementada cabe destacar que los items de visibilidad e interoperabilidad de criterios REBIUN responden a

SciELO Preprints - Este documento es un preprint y su situación actual está disponible en: https://doi.org/10.1590/SciELOPreprints.8124

estándares de la industria editorial y estos deberían articularse con criterios SEO para nutrir un feedback completo en el mundo digital.

PrePrints

Una síntesis del análisis del relevamiento de repositorios PrePrint en base a las plantillas REBIUN muestra los siguientes puntos:

- Cumplimiento de Criterios: Los repositorios evaluados muestran un compromiso significativo con la visibilidad y la accesibilidad de los preprints. Por ejemplo, arXiv y bioRxiv demuestran una alta visibilidad al tener nombres normalizados y URLs amigables, lo que facilita su ubicación por parte de investigadores interesados. Sin embargo, hay áreas donde todos los repositorios tienen margen de mejora. Por ejemplo, la presencia en directorios y recolectores nacionales e internacionales es limitada en la mayoría de los casos, lo que puede afectar la visibilidad y el acceso a los preprints.
- Diversidad Tecnológica: Existe una diversidad en las tecnologías utilizadas por los repositorios, desde desarrollos propios hasta plataformas como OPS y OSF. Por ejemplo, ChinaXiv utiliza una tecnología de desarrollo propio, lo que refleja la adaptabilidad y variedad de enfoques para la gestión de preprints.
- Visibilidad y Accesibilidad: Aunque la mayoría de los repositorios tienen una buena visibilidad en términos de nombres normalizados y URLs amigables, la presencia en directorios y recolectores podría mejorar para aumentar la accesibilidad. Por ejemplo, SciELO PrePrint tiene una presencia limitada en directorios nacionales e internacionales, lo que puede dificultar su descubrimiento por parte de la comunidad científica.
- Políticas y Metadatos: Los repositorios muestran una variedad en la claridad y exhaustividad de sus políticas y la disponibilidad de metadatos. Por ejemplo, arXiv y bioRxiv tienen políticas claras sobre el acceso abierto y la reutilización de metadatos, lo que facilita su interoperabilidad y reutilización por parte de otros sistemas. SciELO presenta políticas escasas, con errores ortográficas y poco desarrollo en su portal específico. Remite a desarrollos políticos el portal general.
- Interoperabilidad: Aunque muchos repositorios cumplen con los requisitos mínimos de interoperabilidad, hay áreas donde podrían mejorar. Por ejemplo, la marcación de registros eliminados y la entrega progresiva de registros a través de lotes son aspectos que podrían fortalecerse en todos los repositorios evaluados, excepto en los que utilizan OPS como SciELO Preprints.

En conclusión, si bien los repositorios de preprints muestran áreas de fortaleza, como la visibilidad y el compromiso con el acceso abierto, también existen oportunidades de mejora en términos de interoperabilidad y presencia en directorios y recolectores. Al abordar estos desafíos, los repositorios pueden desempeñar un papel más efectivo en la difusión de la investigación y el avance del conocimiento científico.

Ahora bien, los relevamientos SEO de cada dominio proporcionaron información valiosa sobre el rendimiento y la visibilidad en línea de cada repositorio, incluyendo su autoridad, tráfico orgánico, backlinks, posición en los resultados de búsqueda y palabras clave relevantes.

- Authority Score: En cuanto al índice más relevante, arxiv.org se destaca en el posicionamiento al tener 68. Sin embargo, muy de cerca están con 64 osf.io/preprints y con 63 preprints.scielo.org. Con 51 puntos está biorxiv.org y muy por detrás chinaxiv.org con 6 puntos.

SciELO Preprints - Este documento es un preprint y su situación actual está disponible en: https://doi.org/10.1590/SciELOPreprints.8124

- Tráfico orgánico: Gran parte del Authority Score está influenciado por el tráfico orgánico que tiene el sitio; así se muestran 1.3 M de visitas para arxiv.org, 99.4 K para biorxiv.org, 83.9 K para osf.io/preprints 18.6 K para preprints.scielo.org, y 4 para chinaxiv.org.
- Backlinks: La mayor parte de backlinks se lo lleva también arxiv.org con 575.8 M, seguido por biorxiv.org tiene 27.8 M, luego osf.io/preprints con 3.4 M, chinaxiv.org 820.3 K y preprints.scielo.org 42.7 K.
- Palabras claves orgánicas: La intención Informativa es la principal en todos los portales. Para arxiv.org y biorxiv.org prima el nombre del repositorio como principal palabra clave, mientras que para los demás se les asigna palabras claves referenciales a artículos publicados como principales. Solo para osf.io/preprints hay una palabra clave que refiere a uno de los preprints que componen el ecosistema de OSF.
- Competidores orgánicos: Para arxiv.org y biorxiv.org prevalecen como competidores orgánicos portales o editoriales de contenido académico como: plos.org, elifesciences.org, paperswithcode.com, openreview.net y aps.org. Para preprints.scielo.org se encuentra dentro de la competencia sitios institucionales como: sld.cu, ucv.ve y scielo.org.mx. También aparece sitios con poco valor de contenido para el sector como goya.es. Finalmente, para osf.io/preprints la principal competencia son sus propios sitios y journals.sagepub.com. chinaxiv.org solo tiene referencias de sitios locales sin relación directa con el sector.

Correlación entre los Resultados de las Plantillas REBIUN y SemRush

Se observa una correlación entre los resultados obtenidos a través de las plantillas REBIUN y los datos de SemRush. Los repositorios que cumplen con los criterios de calidad y accesibilidad tienden a mostrar una mayor autoridad en línea y mejor rendimiento SEO. Sin embargo, se identificaron algunas discrepancias, lo que sugiere que otros factores también pueden influir en el rendimiento en línea de un repositorio.

- Authority Score y Cumplimiento de Criterios REBIUN: Existe una correlación positiva entre el Authority Score y el cumplimiento de criterios REBIUN en términos de visibilidad y accesibilidad de los preprints. Los repositorios con mayor Authority Score tienden a mostrar un compromiso significativo con la visibilidad y accesibilidad, como se evidencia en el caso de **arxiv.org** que tiene altos puntajes tanto en SEO como en cumplimiento de criterios REBIUN. Excepción de esto son biorxiv.org y chinaxiv.org. El primero ya que su análisis REBIUN es bajo mientras que su Authority Score es alto. El segundo es la situación inversa, su análisis REBIUN alto mientras que su Authority Score es bajo. biorxiv.org, sin embargo, puede explicar su alta presencia en la web con una API potente que tal vez los criterios REBIUN no lleguen a compensar con una interoperabilidad pensada solo desde OAI y XML.
  preprints.scielo.org es un caso ejemplar con alto posicionamiento en SEO como en criterios REBIUN. El primero brindado, en parte, por el ecosistema cautivo que el portal tiene con las revistas participantes y, también, por la consolidación de la organización a nivel regional. Respecto a criterios REBIUN, utilizar OPS para su infraestructura posibilita cumplir con gran parte de lo solicitado. Un mayor cuidado del contenido y una dedicación más profunda a la interfaz de usuario podrían aportar al portal un mayor impacto en la comunidad. Finalmente, el ecosistema de servidores preprints OSF presenta un impacto en indicadores SEO considerables debido, en gran parte, a la

SciELO Preprints - Este documento es un preprint y su situación actual está disponible en: https://doi.org/10.1590/SciELOPreprints.8124

implementación de la plataforma en diferentes portales socios que utilizan la infraestructura, como africarxiv y psyarxiv.

- Tráfico Orgánico, Backlinks y Visibilidad REBIUN: El tráfico orgánico refleja la cantidad de visitas que recibe un sitio desde los resultados de búsqueda orgánica. Los repositorios con mayor tráfico orgánico, como arXiv y bioRxiv, pueden estar mejor posicionados en los resultados de búsqueda, lo que contribuye a su visibilidad. Esta correlación sugiere que un mayor tráfico orgánico puede estar asociado con una mejor visibilidad y accesibilidad de los preprints. Ahora bien, luego del análisis los criterios REBIUN parecen estar más centrados en estándares editoriales de implementación internacional que a un estudio detenido para aumentar el tráfico a un sitio, que son los principales objetivos del análisis SEO. Sin embargo, esto no quita la posibilidad de realizar ciertas deducciones:
    - El tráfico web y la nomenclatura clara de un sitio parecen ser un buen indicador.
    - Adicionalmente, la cantidad de backlinks parece estar directamente relacionada con el tráfico web, excepto en el caso de chinaxiv.org y preprints.scielo.org. Sin embargo, los backlinks de chinaxiv.org parecen ser de muy poco autoridad y calidad, por lo que no se refleja ni en la autoridad del dominio ni tráfico orgánico.
    - Otra cuestión fundamental relacionada entre el tráfico web y los backlinks es la trayectoria del portal, por lo que en los sitios con mayor cantidad de años online, los links de referencias son mayores y también el tráfico.
- Competidores y palabras claves: la gestión de metadatos es el principal criterio editorial que impactará en los indicadores SEO de competidores y palabras claves. Según el relevamiento REBIUN los sitios que cumplen con mayores criterios de metadatos son arxiv.org, preprints.scielo.org y chinaxiv.org, mientras que biorxiv.org y osf.io/preprints son los más escasos. Sin embargo, la gestión de palabras claves y competidores relevantes que logra, principalmente biorxiv.org sugiere que tal vez al criterio puramente editorial de utilizar etiquetas Dublin Core que propone REBIUN para evaluar la gestión de metadatos puedan a ampliarse según la experiencia de estos dos portales mencionados que proponen una interoperabilidad basada en el desarrollo de una rebusta API que incluya JSON más que solamente en la interoperabilidad XML o etiquetas HTML.

preprints.latarxiv.org: Producto mínimo viable

Basándonos en los resultados recopilados y analizados anteriormente, hemos avanzado en la creación de un Producto Mínimo Viable (MVP) disponible en preprints.latarxiv.org. Este MVP representa un primer paso crucial en el establecimiento de una infraestructura sólida para la difusión de preprints y datos de investigación en el ámbito de habla hispana.

Objetivos del MVP:

1. **Fomentar la visibilidad:** El MVP se diseñó para mejorar la visibilidad de la investigación en la comunidad científica de habla hispana, proporcionando un espacio centralizado para el acceso a preprints y datos de investigación.

SciELO Preprints - Este documento es un preprint y su situación actual está disponible en: https://doi.org/10.1590/SciELOPreprints.8124

2. **Promover la accesibilidad:** Buscamos hacer que la investigación sea más accesible para una audiencia amplia, incluyendo investigadores, académicos, estudiantes y el público en general interesado en la ciencia y la investigación.

3. **Establecer estándares de calidad:** A través del MVP, nos comprometemos a mantener altos estándares de calidad en términos de metadatos, políticas de acceso abierto y facilidad de uso para los usuarios.

Características clave del MVP:

1. **Visibilidad y Accesibilidad:** Se prioriza una interfaz de usuario intuitiva y fácil de navegar, que facilita la búsqueda, el descubrimiento y la descarga de preprints. Se utiliza OPS (Open Preprint Systems) que posibilita una gestión integral de la publicación además de tener integraciones básicas como DOI con Crossref, entre otros.

   Para cumplimentar este punto, se ha personalizado la interfaz de OPS con una hoja de estilo CSS que muestra a la plataforma con una interfaz moderna supliendo a la que tiene por defecto el sistema. Se han tomado herramientas de otros sitios como: un botón de feedback, un formulario de contacto incrustado, un formulario de suscripción a newsletter, integraciones con estadísticas (se harán visibles en el sitio mediante Data Studio cuando haya datos significativos), entre otros.

   Si bien la URL para el preprint es amigable, inicialmente se quitó la extensión *index.php* característica del software PKP para mejorar la rápida lectura humana. Incluso, las primeras publicaciones incluyen una personalización de la URL del artículo. Sin embargo, luego se decidió rever esta acción mostrando nuevamente el *index.php* en la URL y manteniendo el ID de los preprints en las URL finales para mejorar la rápida indexación de lo publicado en Google Scholar según las indicaciones de la documentación de PKP que indica el formato (PKP Docs, s. f.): https://SITENAME.COM/index.php/JOURNALABBREVIATION/article/view/SUBMISSION#

   Vale destacar que PKP señala que no está firmemente probada esta indicación para OPS y OMP como si para OJS, pero vale la pena avanzar en esa dirección al menos mientras el MVP esté en desarrollo. Con la URL personalizada, el primer artículo publicado (Pitones, 2024) no se indexó en Google Scholar durante los primeros 28 días de publicado, al día 28 se actualizó la URL con lo indicado y se procedió a actualizar los metadatos del DOI correctamente. La falta de indexación inicial pudo deberse a dos cuestiones específicas:

   - Relevamiento inicial del portal, que suele demorar entre 6 y 9 meses, si bien no hay documentación clara al respecto de los tiempos (*Google Scholar Help*, 2024; Scholastica, 2024)
   - Demora adicional por la personalización de la URL. En comparación con una carga adicional que hizo el autor en https://www.scienceopen.com, éste se indexó en Google Scholar rápidamente; sin embargo, los metadatos que el servidor preprint https://www.scienceopen.com/ envío a Crossref y se pueden relevar en ORCID no se corresponden con una tipología de PrePrint sino de Journal Article. Esta característica para el DOI https://doi.org/10.14293/pr2199.000815.v1 que está indicado como PrePrint en la web pero Journal Article en metadatos es una práctica que puede influenciar en la indexación de Google Scholar, sin embargo

SciELO Preprints - Este documento es un preprint y su situación actual está disponible en: https://doi.org/10.1590/SciELOPreprints.8124

también es una práctica descuidada por parte de https://www.scienceopen.com que dificulta posibles futuras relaciones entre las versiones del artículo (Crossref, 2024) y puede perjudicar al autor al no identificar correctamente a un objeto digital.

2. **Políticas y Metadatos:** Se cuidan los estándares de calidad y accesibilidad establecidos por iniciativas como REBIUN, asegurando que nuestros usuarios tengan una experiencia óptima al interactuar con nuestra plataforma. Se desarrollaron políticas claras con una paginación amigable para el usuario, accesibles tanto desde el menú principal como desde el footer. Se han incluido tutoriales de carga, templates normalizadores y un sistema de financiamiento mediante membresía a revistas que deseen un espacio preprint para sus publicaciones. Las normas editoriales y licencias son variables y se adaptan al entorno de publicación sin perder, sin embargo, el estricto control de la coherencia interna sobre la opción elegida.

   La gestión de metadatos es facilitada por OPS con la inserción de etiquetas Dublin Core y Highwire Press Tags, principalmente en una interoperabilidad signada por OAI y XML. A futuro, se podrá desarrollar una API robusta en caso de querer ampliar la interoperabilidad.

3. **Promoción activa:** Se llevarán a cabo estrategias de promoción activa para aumentar la adopción y el conocimiento de Preprints.LatArXiv.org dentro de la comunidad científica de habla hispana, así como en el ámbito internacional. Las mismas incluyen backlinks, presentaciones, difusión por mail, Google ads, entre otros.

4. **Feedback y mejora continua:** Estableceremos canales de retroalimentación para recibir comentarios de los usuarios y así poder realizar mejoras continuas en la plataforma, adaptándonos a las necesidades cambiantes de la comunidad. Se planea desarrollar un especio de comentarios específicos para los artículos además de las ya implementadas herramientas de hypothes.is.

Datos

El segundo componente del ecosistema de LatArXiv es el repositorio de datos. Sobre este, al momento, solo hemos desarrollado la evaluación previa y no se ha desarrollado el MVP como si está plasmado en preprints. Una síntesis del relevamiento de repositorios de datos en base a criterios REBIUN muestra los siguientes puntos:

- Cumplimiento de Criterios: En general, los repositorios evaluados muestran un compromiso significativo con la visibilidad y accesibilidad de los datos. Por ejemplo, Repositorio ANID y Repositorio Institucional de la Universidad Nacional de Colombia cumplen con una amplia gama de criterios, desde la presencia en directorios hasta la disponibilidad de metadatos. Sin embargo, aún hay áreas de mejora, como la presencia en recolectores nacionales e internacionales y la claridad de las políticas.
- Diversidad Tecnológica: Se observa una diversidad en las tecnologías utilizadas por los repositorios, desde plataformas propias como Dataverse hasta sistemas como DSpace. Esta diversidad refleja la adaptabilidad y variedad de enfoques para la gestión de datos.
- Visibilidad y Accesibilidad: Aunque la mayoría de los repositorios tienen una buena visibilidad en términos de nombres normalizados y URLs amigables, la presencia en directorios y recolectores podría mejorar para aumentar la

SciELO Preprints - Este documento es un preprint y su situación actual está disponible en: https://doi.org/10.1590/SciELOPreprints.8124

accesibilidad. Por ejemplo, SciELO Data muestra una presencia limitada en directorios, lo que puede dificultar su descubrimiento por parte de la comunidad investigadora.

- Políticas y Metadatos: Existe una variedad en la claridad y exhaustividad de las políticas y la disponibilidad de metadatos entre los repositorios evaluados. Algunos, como Repositorio ANID, tienen políticas claras sobre acceso abierto y preservación, mientras que otros, como SciELO Data, presentan políticas escasas y poco desarrolladas.
- Interoperabilidad: Aunque muchos repositorios cumplen con los requisitos mínimos de interoperabilidad, hay áreas donde podrían mejorar, como la marcación de registros eliminados y la entrega progresiva de registros a través de lotes. La adopción de estándares como OAI-PMH puede fortalecer la interoperabilidad entre los repositorios.

Por su parte, los relevamientos SEO de cada dominio de repositorio de datos proporcionaron la siguiente información:

- Authority Score: en primer lugar datosabiertos.unam.mx con 83, seguido por los demás sitios: sedici.unlp.edu.ar 65, data.scielo.org 63, repositorio.unal.edu.co 61 y repositorio.anid.cl 42.
- Tráfico orgánico: en primer lugar sedici.unlp.edu.ar con 268.7 K, seguido por los demás sitios: repositorio.unal.edu.co 152.8 K, datosabiertos.unam.mx con 7.5 K, data.scielo.org 1.9 K y repositorio.anid.cl 4.
- Backlinks: La mayor parte de backlinks se lo lleva también sedici.unlp.edu.ar con 3.3 M, seguido por repositorio.unal.edu.co que tiene 556.9 K, luego repositorio.anid.cl con 408.3 K, datosabiertos.unam.mx 6.4 K y data.scielo.org 1.2 K.
- Palabras claves orgánicas: La intención Informativa es la principal en todos los portales de depósitos de datos, aunque la intención transaccional cobra importancia para este tipo específico de sitios. En todos los portales, las palabras claves que llevan al portal es por el contenido y nos por el portal en sí.
- Competidores orgánicos: Para sedici.unlp.edu.ar aparece competidores como humanidades.com y uba.ar, para repositorio.unal.edu.co ucsp.edu.pe y areandina.edu.co. En cuanto a repositorio.anid.cl tenemos como competidores a educalt.com e ibero.edu.co entre los primeros, mientras que para datosabiertos.unam.mx tenemos a concepto.de y vademécum.es. Finalmente, para data.scielo.org seguimos teniendo al igual que en preprints a sld.cu, ucv.ve y scielo.org.mx, entre otros.

Correlación entre los Resultados de las Plantillas REBIUN y SemRush

Para analizar la correlación entre los resultados obtenidos a través de las plantillas REBIUN y los datos de SemRush para los repositorios de datos, podemos considerar los siguientes puntos:

1. **Authority Score y Cumplimiento de Criterios REBIUN**: Existe una correlación positiva entre el Authority Score y el cumplimiento de criterios REBIUN en términos de visibilidad y accesibilidad de los datos. Los repositorios con un Authority Score más alto tienden a cumplir con una amplia gama de criterios REBIUN, como la presencia en directorios, la disponibilidad de metadatos y la claridad de las políticas. Esto sugiere que los repositorios bien

SciELO Preprints - Este documento es un preprint y su situación actual está disponible en: https://doi.org/10.1590/SciELOPreprints.8124

establecidos y reconocidos en línea también tienden a tener mejores prácticas en términos de gestión de datos, como es el caso de **sedici.unlp.edu.ar.**

2. **Tráfico Orgánico y Visibilidad REBIUN**: El tráfico orgánico refleja la cantidad de visitas que recibe un sitio desde los resultados de búsqueda orgánica, lo que puede estar relacionado con su visibilidad en línea. Los repositorios con un mayor tráfico orgánico, como sedici.unlp.edu.ar, pueden estar mejor posicionados en los resultados de búsqueda, lo que contribuye a su visibilidad y accesibilidad. Esto sugiere que una mayor visibilidad en línea puede estar asociada con un mayor tráfico orgánico y viceversa, lo que respalda la importancia de cumplir con los criterios REBIUN para mejorar la visibilidad y accesibilidad de los datos.

3. **Backlinks y Presencia en Directorios y Recolectores REBIUN**: La cantidad y calidad de backlinks pueden estar relacionadas con la presencia en directorios y recolectores, lo que puede influir en la visibilidad y accesibilidad de los datos. Los repositorios con más backlinks, como sedici.unlp.edu.ar, pueden tener una mayor presencia en línea y, por lo tanto, una mayor visibilidad. Esto sugiere que la promoción y la colaboración con otros sitios pueden mejorar la visibilidad y accesibilidad de los datos, aspectos que también son evaluados por los criterios REBIUN.

4. **Competidores Orgánicos y Palabras Claves REBIUN**: La gestión de palabras clave y la presencia de competidores orgánicos pueden estar relacionadas con la relevancia y visibilidad de los datos en línea. Los repositorios que utilizan palabras clave relevantes y tienen competidores orgánicos relacionados con el contenido pueden tener una mejor visibilidad en línea y una mayor atracción de tráfico. Esto sugiere que la gestión efectiva de palabras clave y la identificación de competidores orgánicos pueden mejorar la visibilidad y accesibilidad de los datos, aspectos también considerados por los criterios REBIUN.

En resumen, los resultados obtenidos a través de las plantillas REBIUN y los datos de SemRush proporcionan información valiosa sobre la visibilidad, accesibilidad y rendimiento en línea de los repositorios de datos. Existe una correlación entre estos dos conjuntos de datos, lo que sugiere que cumplir con los criterios REBIUN puede contribuir al éxito en términos de SEO y visibilidad en línea de los repositorios de datos.

SciELO Preprints - Este documento es un preprint y su situación actual está disponible en: https://doi.org/10.1590/SciELOPreprints.8124

Conclusiones

El proyecto LatArXiv ha realizado un exhaustivo análisis de repositorios de preprints y datos de investigación con relevancia para América Latina y España, estableciendo una base sólida para el desarrollo de una nueva infraestructura de acceso abierto en la región de habla hispana. Los resultados obtenidos revelan un panorama inicial con un gran potencial de desarrollo, especialmente en el ámbito de los preprints, mientras que los repositorios de datos presentan una práctica institucional más consolidada.

Preprints: Los repositorios de preprints muestran fortalezas significativas en visibilidad y accesibilidad, cumpliendo con criterios como nombres normalizados y URLs amigables. Sin embargo, se identificaron áreas de mejora en la interoperabilidad y la presencia en directorios y recolectores nacionales e internacionales. El análisis SEO indica que repositorios como arXiv y bioRxiv tienen altos puntajes de Authority Score y tráfico orgánico, aunque otros como ChinaXiv necesitan mejorar su visibilidad en línea. Tal vez la diferenciación entre un factor disciplinar a uno geográfico pueda ser índice de este bajo impacto, sin embargo SciELO con su desarrollo demuestra que portales generales con otros factores de identificación más allá de lo disciplinar pueden alcanzar gran visibilidad y aporte a la comunidad científica.

Repositorios de Datos: Los repositorios de datos evaluados presentan un compromiso significativo con la visibilidad y accesibilidad de los datos, aunque también se destacan áreas para mejorar, como la claridad de las políticas y la interoperabilidad. Repositorios como el de la Universidad Nacional de La Plata y el de la Universidad Nacional de Colombia muestran un alto cumplimiento de criterios REBIUN y buenos resultados en indicadores SEO, lo que sugiere una correlación positiva entre prácticas de gestión de datos y rendimiento en línea.

Correlaciones entre Criterios REBIUN y SEO: Se observa una correlación positiva entre el cumplimiento de criterios REBIUN y el rendimiento SEO. Los repositorios que cumplen con criterios de visibilidad y accesibilidad tienden a tener una mayor autoridad en línea y mejor tráfico orgánico. La gestión efectiva de metadatos y políticas claras también contribuyen a una mejor visibilidad y accesibilidad, reflejándose en un mayor impacto en la comunidad científica. Sin embargo, las discrepancias a esta regla común, como biorxiv.org, sugieren que el desarrollo de herramientas de interoperabilidad modernas basadas en API REST contribuyen considerablemente a la visibilización y utilización del portal, y estos son criterios técnicos que no abordan las plantillas REBIUN.

Desarrollo de MVP: El desarrollo del Producto Mínimo Viable (MVP) de LatArXiv para preprints ha sido un paso crucial. Este MVP prioriza la visibilidad y accesibilidad, con una interfaz de usuario intuitiva y políticas claras, cumpliendo con los estándares de calidad establecidos por REBIUN y complementando con estrategias SEO. Este enfoque permite que LatArXiv se posicione como una plataforma robusta y accesible para la difusión de preprints y, eventualmente, de datos de investigación.

Recomendaciones: Para fortalecer la infraestructura de acceso abierto en la región, se recomienda:

- Mejorar la interoperabilidad y la presencia en directorios y recolectores.
- Implementar estrategias SEO más efectivas para aumentar la visibilidad y el tráfico orgánico.

SciELO Preprints - Este documento es un preprint y su situación actual está disponible en: https://doi.org/10.1590/SciELOPreprints.8124

- Desarrollar políticas claras y exhaustivas que promuevan la accesibilidad y la reutilización de los datos.
- Continuar el desarrollo del repositorio de datos siguiendo los mismos estándares de calidad aplicados al MVP de preprints.

LatArXiv tiene el potencial de convertirse en un referente para la comunidad científica de habla hispana, promoviendo la colaboración y el intercambio de conocimiento a través de una infraestructura de acceso abierto sólida y accesible.

Nota

Para la realización de este trabajo final se utilizaron herramientas de IA tanto para el desarrollo y asistencia en códigos R y Python como para la asistencia en redacción de texto. Adicionalmente, en la implementación del MVP montado en https://preprints.latarxiv.org/ se recibió asistencia y apoyo de personal de https://paideiastudio.net/ con conocimientos específicos en CSS para la personalización y PHP para soluciones de errores relacionados a la versión de OPS para que el montaje final sea funcional. Finalmente, la gestión integral del proyecto, montaje, redacción de contenido, estructura y gran parte del código personalizado de OPS estuvo a cargo del autor mencionado en la carátula y que hace la presentación del trabajo.

Conflicto de Interés

El autor es director de Paideia Studio, entidad editora de LatArXiv.

Conflict of Interest

The author is director of Paideia Studio, publisher of LatArXiv.

Referencias

Barrueco Cruz, J. M., Rico-Castro, P., & Bonora Eve, L. V. (2021, mayo 6). *Guía para la evaluación de repositorios institucionales de Investigación* [Text]. FECYT. https://www.fecyt.es/es/publicacion/guia-para-la-evaluacion-de-repositorios-institucionales-de-investigacion

Crossref. (2024). *Version control, corrections, and retractions* [Website]. Crossref. https://www.crossref.org/documentation/crossmark/version-control-corrections-and-retractions/

Giles, J. (2003). Preprint server seeks way to halt plagiarists. *Nature*, *426*(6962), 7-8.

Google Scholar Help. (2024). https://scholar.google.com/intl/en/scholar/inclusion.html#overview

Kaiser, J. (2017). The preprint dilemma. *Science*, *357*(6358), 1344-1349. https://doi.org/10.1126/science.357.6358.1344

SciELO Preprints - Este documento es un preprint y su situación actual está disponible en: https://doi.org/10.1590/SciELOPreprints.8124

Kwon, D. (2020). How swamped preprint servers are blocking bad coronavirus research. *Nature*, *581*(7807), 130-131. https://doi.org/10.1038/d41586-020-01394-6

Pitones, L. E. M. (2024). *Auditorias SEO en plataformas de automatización de marketing y su relación con el posicionamiento orgánico en SERPS*. LatArXiv. https://doi.org/10.62059/LatArXiv.preprints.29

PKP Docs. (s. f.). *Google Scholar Indexing*. PKP Docs. Recuperado 14 de mayo de 2024, de https://docs.pkp.sfu.ca/google-scholar/en/

Scholastica. (2024). *How does Google Scholar indexing work? - Scholastica Help Desk*. https://help.scholasticahq.com/article/214-how-does-google-scholar-indexing-work-on-scholastica

van Schalkwyk, F., & Dudek, J. (2022). Reporting preprints in the media during the COVID-19 pandemic. *Public Understanding of Science (Bristol, England)*, *31*(5), 608-616. https://doi.org/10.1177/09636625221077392

Este preprint fue presentado bajo las siguientes condiciones:

- Los autores declaran que son conscientes de que son los únicos responsables del contenido del preprint y que el depósito en SciELO Preprints no significa ningún compromiso por parte de SciELO, excepto su preservación y difusión.
- Los autores declaran que se obtuvieron los términos necesarios del consentimiento libre e informado de los participantes o pacientes en la investigación y se describen en el manuscrito, cuando corresponde.
- Los autores declaran que la preparación del manuscrito siguió las normas éticas de comunicación científica.
- Los autores declaran que los datos, las aplicaciones y otros contenidos subyacentes al manuscrito están referenciados.
- El manuscrito depositado está en formato PDF.
- Los autores declaran que la investigación que dio origen al manuscrito siguió buenas prácticas éticas y que las aprobaciones necesarias de los comités de ética de investigación, cuando corresponda, se describen en el manuscrito.
- Los autores declaran que una vez que un manuscrito es postado en el servidor SciELO Preprints, sólo puede ser retirado mediante solicitud a la Secretaría Editorial deSciELO Preprints, que publicará un aviso de retracción en su lugar.
- Los autores aceptan que el manuscrito aprobado esté disponible bajo licencia Creative Commons CC-BY.
- El autor que presenta el manuscrito declara que las contribuciones de todos los autores y la declaración de conflicto de intereses se incluyen explícitamente y en secciones específicas del manuscrito.
- Los autores declaran que el manuscrito no fue depositado y/o previamente puesto a disposición en otro servidor de preprints o publicado en una revista.
- Si el manuscrito está siendo evaluado o siendo preparando para su publicación pero aún no ha sido publicado por una revista, los autores declaran que han recibido autorización de la revista para hacer este depósito.
- El autor que envía el manuscrito declara que todos los autores del mismo están de acuerdo con el envío a SciELO Preprints.